DOCUMENT RESUME

ED 477 934 TM 035 031

AUTHOR	MacQuarrie, Duncan
TITLE	Validity Evidence for Washington Assessment of Student Learning (WASL) Performance Standard Cut-Scores for Reading and Mathematics.
PUB DATE	2003-04-00
NOTE	14p.; Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, IL, April 21-25, 2003).
PUB TYPE	Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE	EDRS Price MF01/PC01 Plus Postage.
DESCRIPTORS	Achievement Tests; *Cutting Scores; *Mathematics; Norm Referenced Tests; Performance Factors; *Reading; State Programs; *State Standards; *Testing Programs; *Validity
IDENTIFIERS	Washington; *Washington Assessment of Student Learning

ABSTRACT

This study was designed to contribute to the validity evidence for the Washington Assessment of Student Learning (WASL) by providing additional descriptive data about the performance standards in reading and mathematics at grades 4, 7, and 10. After the realignment of norm-referenced tests, large numbers of students taking the WASL had corresponding norm-referenced scores from the previous year. It was possible to match samples for both sets of tests. Students' performance on the norm-referenced tests consistently showed mathematics performance to be slightly higher than reading performance at all grade levels, and performance across grade levels for both reading and mathematics was quite similar. Performance on the standards-based assessments for reading and mathematics, and across grade levels, exhibited marked variations, with mathematics performance consistently lower than corresponding grade level reading performance. Coefficients suggest a moderately strong relationship between performance on the norm-referenced tests and the standards-based assessments given a year later. Equipercentile equating of the distributions from both was developed. In addition, the percentage of students meeting the performance standard was plotted as a function of progressively higher national percentile rank bands. Data and portrayals clearly indicate inconsistencies in the difficulty of performance standards across grade levels and content areas. The lack of vertical comparability for the reading standards at grades 4, 7, and 10 undermines a belief in their reasonableness. Even though they are more consistent, the overall difficulty of the mathematics standards also makes it hard to believe that they are reasonable. The difference between reading and mathematics performance at grades 4 and 10 also makes it difficult to promote these measures as fair. Some of the factors contributing to these problems are discussed. (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
✓ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
D. MacQuarrie
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
1

ED 477 934

Validity Evidence for Washington Assessment of Student Learning (WASL) Performance Standard Cut-Scores for Reading and Mathematics

Duncan MacQuarrie
Tacoma Public Schools

TM035031

Paper presented during the symposium "Do standards-based tests differ fundamentally from norm-referenced tests in practice?" at the Annual Meeting of the American Educational Research Association, Chicago, April 21-25, 2003.

BEST COPY AVAILABLE

Validity Evidence for Washington Assessment of Student Learning (WASL) Performance Standard Cut-Scores for Reading and Mathematics

The performances of schools and districts on the Washington Assessment of Student Achievement (WASL), the state's standards-based assessment, are the primary achievement indicators for the state accountability system. In addition, these same scores are used as the performance indicators in the accountability system required by the No Child Left Behind (NCLB) federal legislation. Critical elements of such standards-based assessments are the performance standards, or cut-scores, that categorize the performance into a limited number of levels. The NCLB requires a minimum of three levels and labels them "basic," "proficient," and "advanced." The categories of "proficient" and "advanced" are considered acceptable levels of achievement in these new accountability systems. Therefore, the validity of these classifications, and the inferences about students and schools that are based on them, are of great importance.

The cut-scores are typically arrived at through a standard setting procedures based on judgments. Such decisions are made by panels of judges, primarily educators having knowledge of the curriculum standards from which the test content is derived and experience teaching students at the grade level being tested. In the current climate of high stakes accountability, any number of such performance standards for state tests are perceived as unreasonably difficult. Such doubts about the fairness of these performance standards raise questions of the validity of the interpretations about students and schools that are based on them.

The WASL was phased in over three consecutive years beginning with 4th grade in the spring of 1997. In the initial years these assessments were voluntary for schools and districts. However, at each grade level, over ninety percent of the students in the state participated during these voluntary years. The 4th grade assessment, voluntary in the spring of 1997, became mandatory in the spring of 1998. The 7th grade assessment was instituted as a voluntary assessment in the spring of 1998 and did not become mandatory

until the spring of 2001. The 10th grade assessment was the last to be developed and first appeared as a voluntary program in the spring of 1999. This component also became mandatory in the spring of 2001. The performance standards (cut-scores) for these assessments were established during the summer immediately following their initial administration.

The Washington State Assessment Program also includes three grade levels of norm-referenced tests. The Iowa Tests of Basic Skills (ITBS) is administered in the spring at 3rd and 6th grades and the Iowa Tests of Educational Development (ITED) at 9th grade. These assessments represent a hold over from the prior state assessment program and used to be administered in grades 4, 8, and 10. However, with the institution of the standards-based assessments in grades 4, 7, and 10, it was decided a better alignment would be to place the "basic skills" assessments in the years prior to the standards-based tests. These placements occurred first at the elementary level in the spring of 1999 and in the following year at the secondary level.

The study reported here was designed to contribute to the validity evidence for the WASL by providing additional descriptive data about the performance standards in reading and mathematics at 4th, 7th, and 10th grades. After the realignment of the norm-referenced tests large numbers of students taking the WASL had corresponding norm-references test scores from the previous year. The first such cohort with both the prior year's norm-referenced test scores and the corresponding standards-based scores occurred in the spring of 2000 at 4th grade. In the subsequent spring of 2001 such cohorts first occurred at 7th and 10th grades. Table 1 shows the percent of students meeting the state performance standard in grades 4, 7 and 10 for reading and mathematics for all students and for the matched sets of students having norm-referenced test scores from the prior year for each of these cohorts. These matched samples included only students having valid scores for reading and mathematics on both the standards-based assessments and the norm-referenced tests. Table 2 shows the ITBS or ITED National Percentile Rank (NPR) equivalent of the mean scale scores in grades 3, 6, and 9 for all students and the corresponding matched samples represented in Table 1.

Students' performance on the norm-referenced tests (Table 2) consistently shows mathematics performance to be slightly higher than reading performance at all grade levels. In addition, the performance across grade levels for both reading and mathematics was quite similar. Performance on the standards-based assessments (Table 1) for reading and mathematics, and across grade levels, exhibited marked variations. Performance on the mathematics assessments is uniformly lower than the corresponding grade level reading performance. Math performance is highest at the elementary level and lowest at the middle level. Reading performance is much higher than math at grades 4 and 10. Although the 7th grade reading performance is still higher than math at that level, it is markedly below that for reading at grades 4 and 10. These patterns raise concerns about the reasonableness of the performance standards for the standards-based assessments, particularly give the corresponding stability in the norm-referenced test data.

Table 4 shows the correlation coefficients for the norm-referenced and standards-based reading and math pairs for the three grade levels for the different matched samples. These correlation coefficients remained quite consistent across years with the exception of that for reading between the ITBS reading at 3rd grade in 2001 and the WASL reading at 4th grade in 2002. These coefficients suggest a moderately strong relationship between the performance on the norm-referenced tests and the standards-based assessments given a year later. Based on the size of these coefficients, two additional analyses were conducted.

First, equipercentile equating of the standards-based assessment distributions and the corresponding norm-referenced distributions were developed. Table 5 gives the estimated NPR for the equivalent standards-based assessment cut-score at the performance standard. When expressed as NPRs it is clear that the reading cut-scores at the standard in 4th and 10th grades are at the lower end of what would be considered the normal or average range of traditional norm-referenced test performance. The math cut-score at 4th grade appears to be at the upper end of the normal range as does that for 7th

grade reading. The math cut-scores for the standards-based assessment at both 7th and 10th grade appear to lie slightly above the normal range.

The second approach to illuminating the relationship between students' prior year norm-referenced test performances and their subsequent standards-based performance involved plotting the percentage of students meeting the performance standard as a function of progressively higher NPR bands ranging from "1-4" to "95-99." Figures 1 through 6 display these relationships. Figures 1 and 2 each show the relationships between the 3rd grade reading (Figure 1) and math (Figure 2) norm-referenced performance and the corresponding standards-based performance at 4th grade. Figure 1 shows that for reading the relationship remained very stable across three consecutive years. Figure 2, for 4th grade math, however shows that the first two years remained almost identical, however for 2002 the percent of students meeting the performance standard was systematically higher for each band except for the two extreme bands. In addition, Figure 2 shows that the percent of students meeting the standard is below 50% until the "60-65 NPR" band is reached. By comparison, for 4th grade reading at the "60-65 NPR" band, over 80% of the students met the standard.

Both Figures 3 and 4 (7th grade reading and math respectively) show slight increases in the percent of students meeting the standard across almost all bands for 2002 compared to 2001. However, the percent of students meeting the standard remains low for both reading and math until the higher bands of the NPR distribution are reached. This is particularly pronounced for 7th grade mathematics.

Finally, Figures 5 and 6 show the relationships between the norm-referenced scores and the standards-based scores for reading and math respectively at 10th grade. The reading function looks very similar to that at 4th grade except there was more growth between 2001 and 2002 than was shown at 4th grade. Math on the other hand shows no growth, actually a slight decline in performance, from 2001 to 2002. In addition, the math function looks much more like that for math at seventh grade except at the higher NPR

bands where slightly larger percents of students met the standard at 10th grade than did for the corresponding bands at 7th grade.

These data and portrayals clearly indicate inconsistencies in the difficulty of the performance standards across grade levels and content areas. The lack of vertical comparability for the reading standards at grades 4, 7, and 10 undermines a belief in their reasonableness. Even though they are more consistent across grade levels, the overall difficulty of the mathematics standards also makes it harder to believe that they are reasonable. The large difference between the reading and math performance at grades 4 and 10 also makes it difficult to promote these accountability measures as fair.

The performance standards for the WASL assessments were set by difference standard setting committees meeting during the summer in three different years (1997, 1998, and 1999). Furthermore, the standard setters were not allowed to have access to impact data during their review process. And finally, the policy board responsible for establishing the performance standards choice to not intervene and moderate the committee recommendations. These factors no doubt contributed in significant ways to produce the results describe in this paper. Much more attention must be paid to the role of policy bodies in the setting of performance standards for these new accountability systems. The work of the judges during the standard setting sessions must be treated as only one source of information about the desired standards. Policy makers must be much better informed about their role in exercising the final judgments about these very important decisions. They must provide the needed moderation required to arrive at performance standards that are perceive as reasonable while at the same time encouraging practitioners to strive for even greater learning for their students.

Washington Data: Norm-Reference & Standards-Based Tests*

Table 1. Standards-Based Tests (SBT) - Percent Met Standard

Year Introduced			Statewide			Matched Sample					
			2000	2001	2002	2000	N	2001	N	2002	N
1997	4th	Reading	65.8	66.1	65.6	71.3	53,092	71.0	57,348	70.0	57,571
		Math	41.8	43.4	51.8	46.6		48.1		56.3	
1998	7th	Reading	41.5	39.8	44.5	NA	NA	44.5	56,430	49.2	58,698
		Math	28.2	27.4	30.4	NA		31.4		34.3	
1999	10th	Reading	59.8	62.4	59.2	NA	NA	72.4	53,372	68.2	55,269
		Math	35.0	38.9	37.3	NA		47.2		44.3	

Table 2. Norm-referenced Tests (NRT) - NPR Equivalent of Mean Scale Score

Year Introduced			Statewide			Matched Sample		
			1999	2000	2001	1999	2000	2001
1999	3th	Reading	55	56	57	53	55	55
		Math	60	63	64	58	59	61
2000	6th	Reading	NA	54	53	NA	55	55
		Math	NA	56	56	NA	57	56
2000	9th	Reading	NA	54	53	NA	59	58
		Math	NA	60	59	NA	65	64

Table 3. Means and Standard Deviations for NRT and SBT Scale Scores - Matched Samples

			Mean			SD		
			99/00	00/01	01/02	99/00	00/01	01/02
4th	NRT	Reading	187.4	188.1	188.3	19.8	19.5	19.3
		Math	188.7	190.2	190.6	18.4	18.5	18.4
	SBT	Reading	409.3	407.6	409.1	18.9	17.9	19.5
		Math	394.9	397.0	403.8	33.7	33.9	33.0
7th	NRT	Reading	NA	230.6	230.5	NA	27.5	27.7
		Math	NA	232.8	232.4	NA	27.6	27.7
	SBT	Reading	NA	396.7	397.2	NA	19.7	19.1
		Math	NA	374.4	379.3	NA	50.3	47.2
10th	NRT	Reading	NA	268.0	267.0	NA	34.3	34.3
		Math	NA	278.1	276.7	NA	35.9	36.1
	SBT	Reading	NA	413.7	411.3	NA	28.8	30.0
		Math	NA	395.8	393.3	NA	40.1	37.2

Table 4. Correlation Coefficients: Prior Year's NRT and Standards-Based Tests

		2000	2001	2002
4th	Reading	.72	.72	.66
	Math	.77	.77	.74
7th	Reading	NA	.76	.75
	Math	NA	.83	.83
10th	Reading	NA	.74	.74
	Math	NA	.80	.80

Table 5. Equipercentile Equating: Estimated NPR Equivalents of the SBT Cut Scores

		2000	2001	2002
4th	Reading	38th	40th	38th
	Math	61st	61st	53rd
7th	Reading	NA	63rd	56th
	Math	NA	72nd	69th
10th	Reading	NA	43rd	45th
	Math	NA	72nd	72nd

*NRT: 3rd & 6th - ITBS; 9th - ITED SBT: 4th, 7th & 10th - Washington Assessment of Student Learning

Figure 1. Percent of Students Scoring in 3rd Grade ITBS Reading NPR Bands Subsequently Meeting 2000, 2001, and 2002 4th Grade WASL Reading Standard

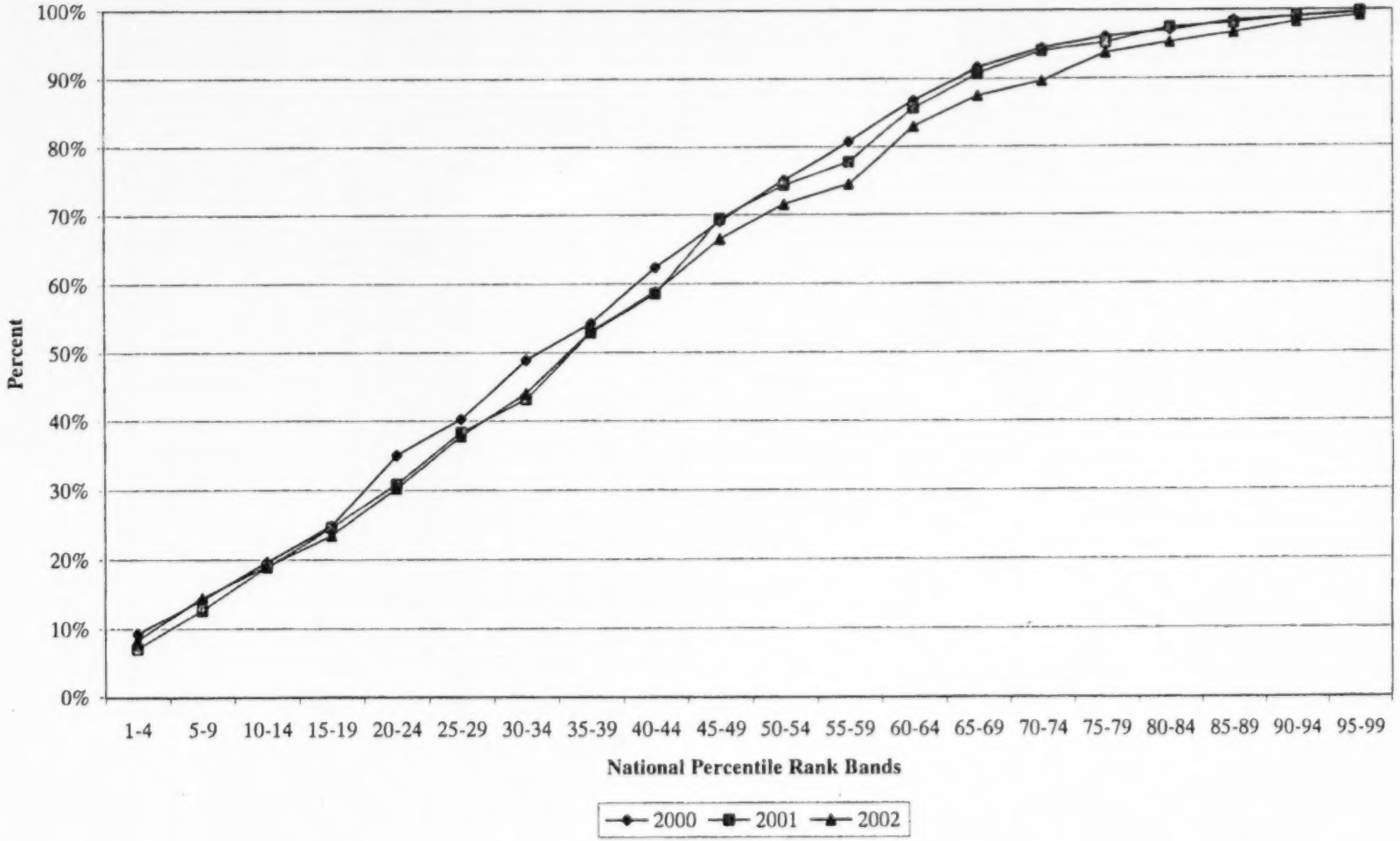


Figure 2. Percent of Students Scoring in 3rd Grade ITBS Math NPR Bands Subsequently Meeting 2000, 2001, and 2002 4th Grade WASL Math Standard

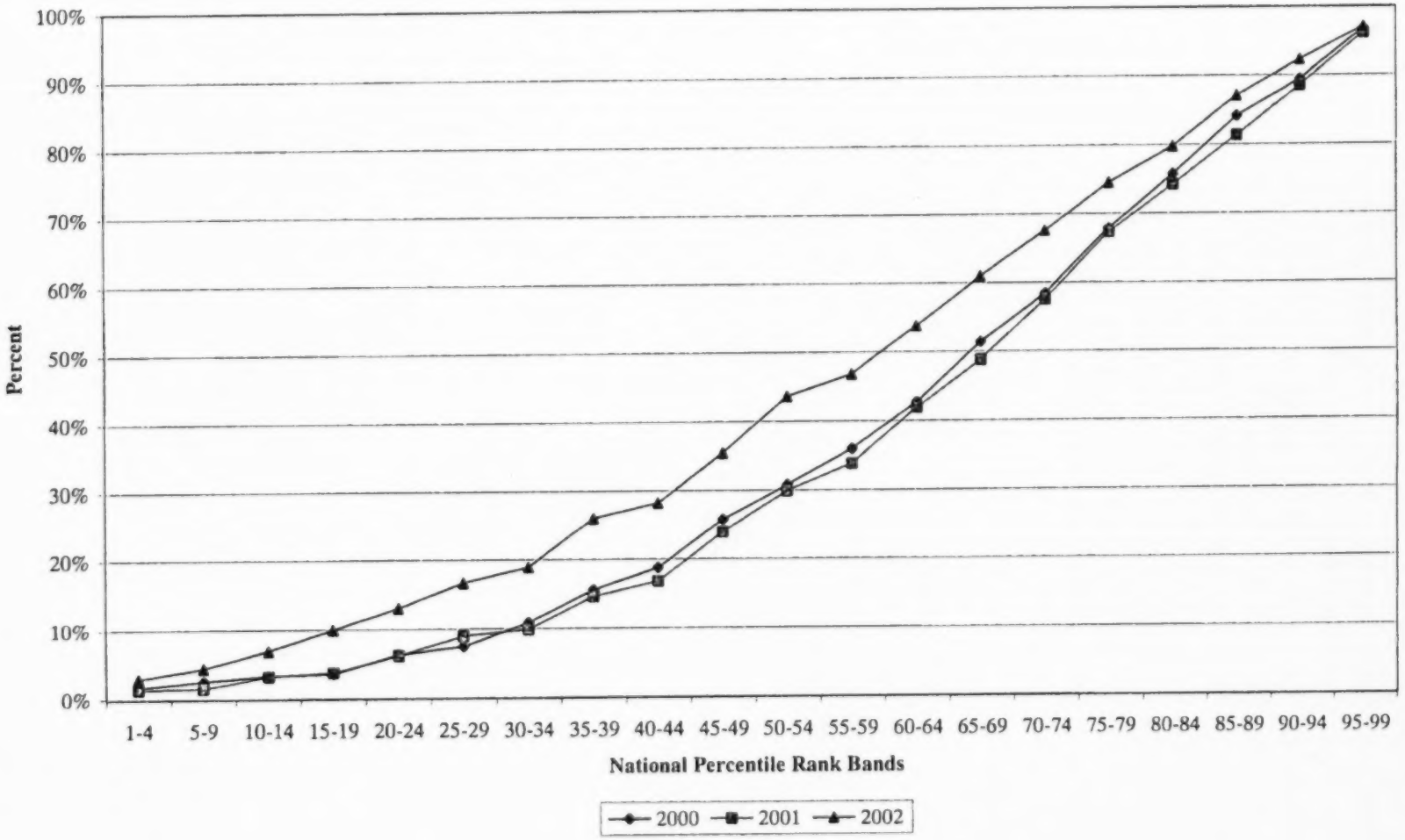


Figure 3. Percent of Students Scoring in 6th Grade ITBS Reading NPR Bands Subsequently Meeting 2001 and 2002 7th Grade WASL Reading Standard

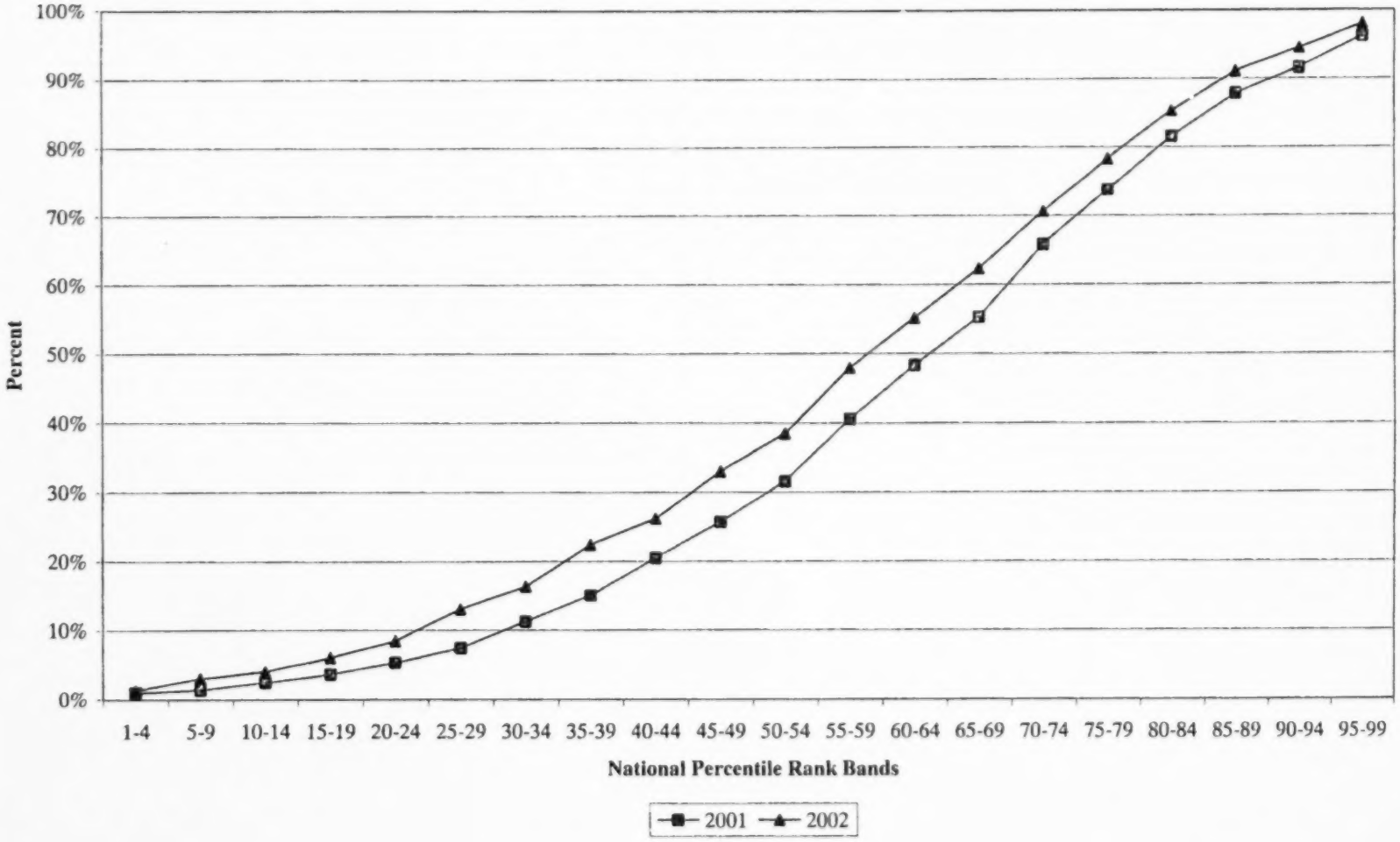


Figure 4. Percent of Students Scoring in 6th Grade ITBS Math NPR Bands Subsequently Meeting 2001 and 2002 7th Grade WASL Math Standard

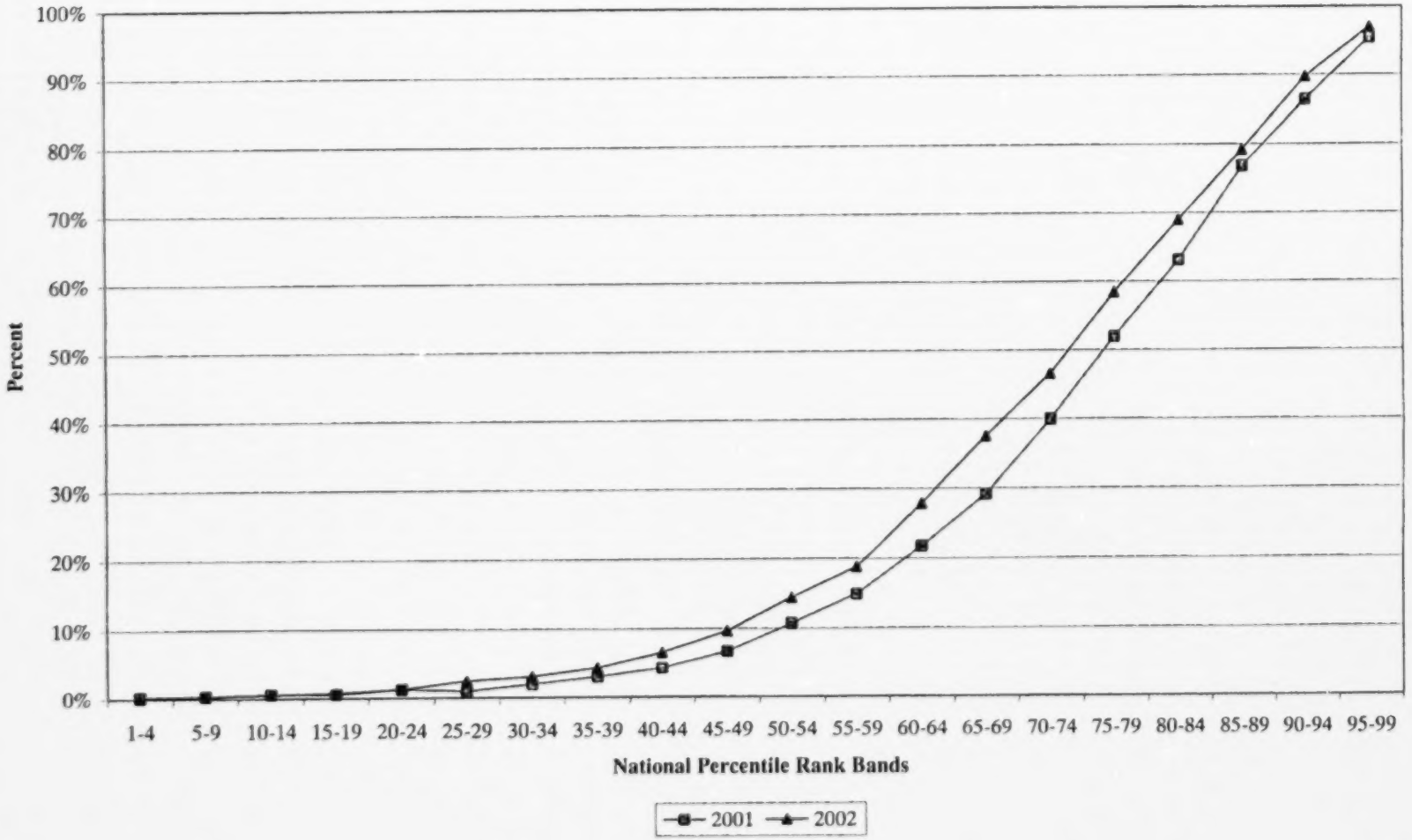


Figure 5. Percent of Students Scoring in 9th Grade ITED Reading NPR Bands Subsequently Meeting 2001 and 2002 10th Grade WASL Reading Standard

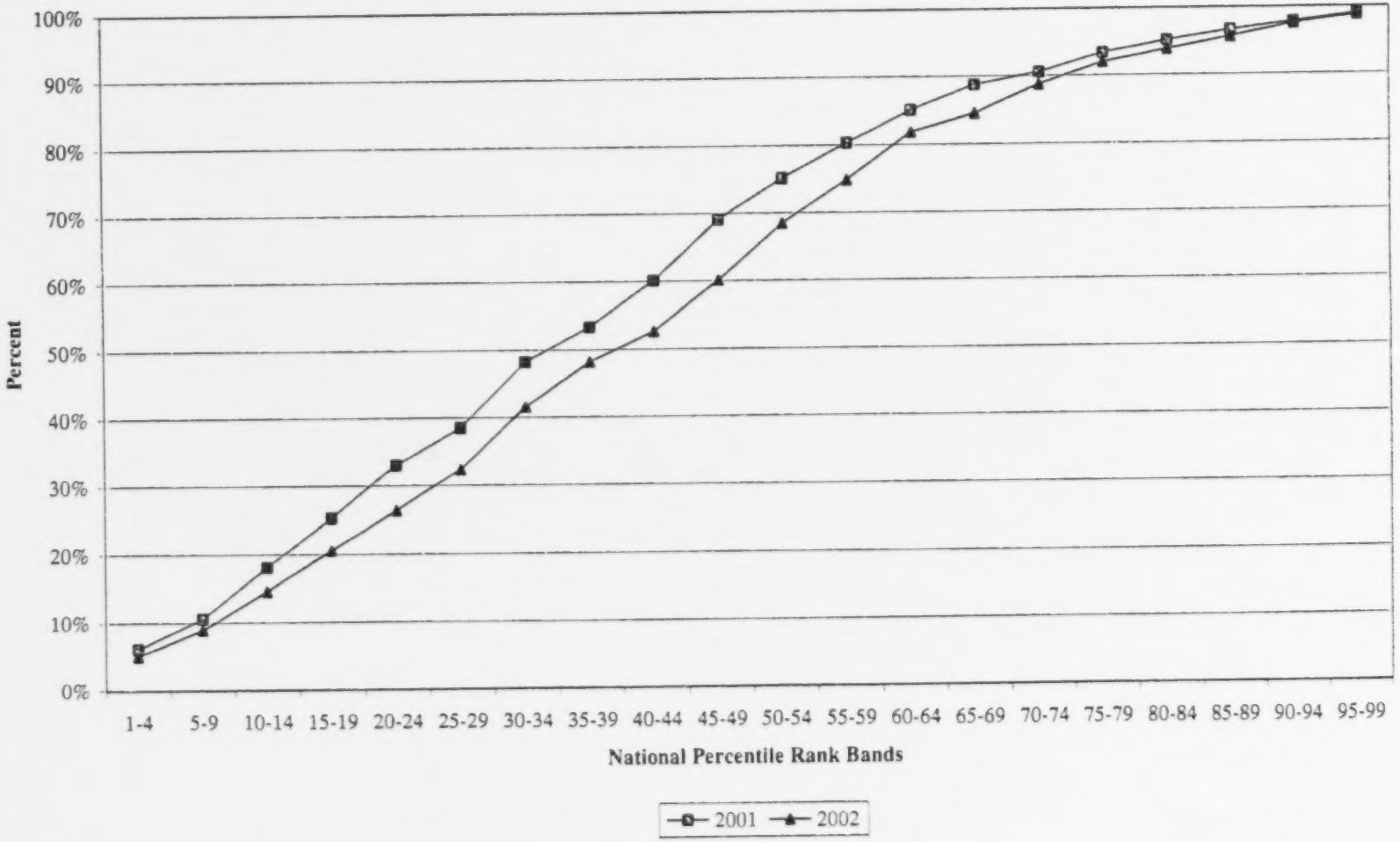
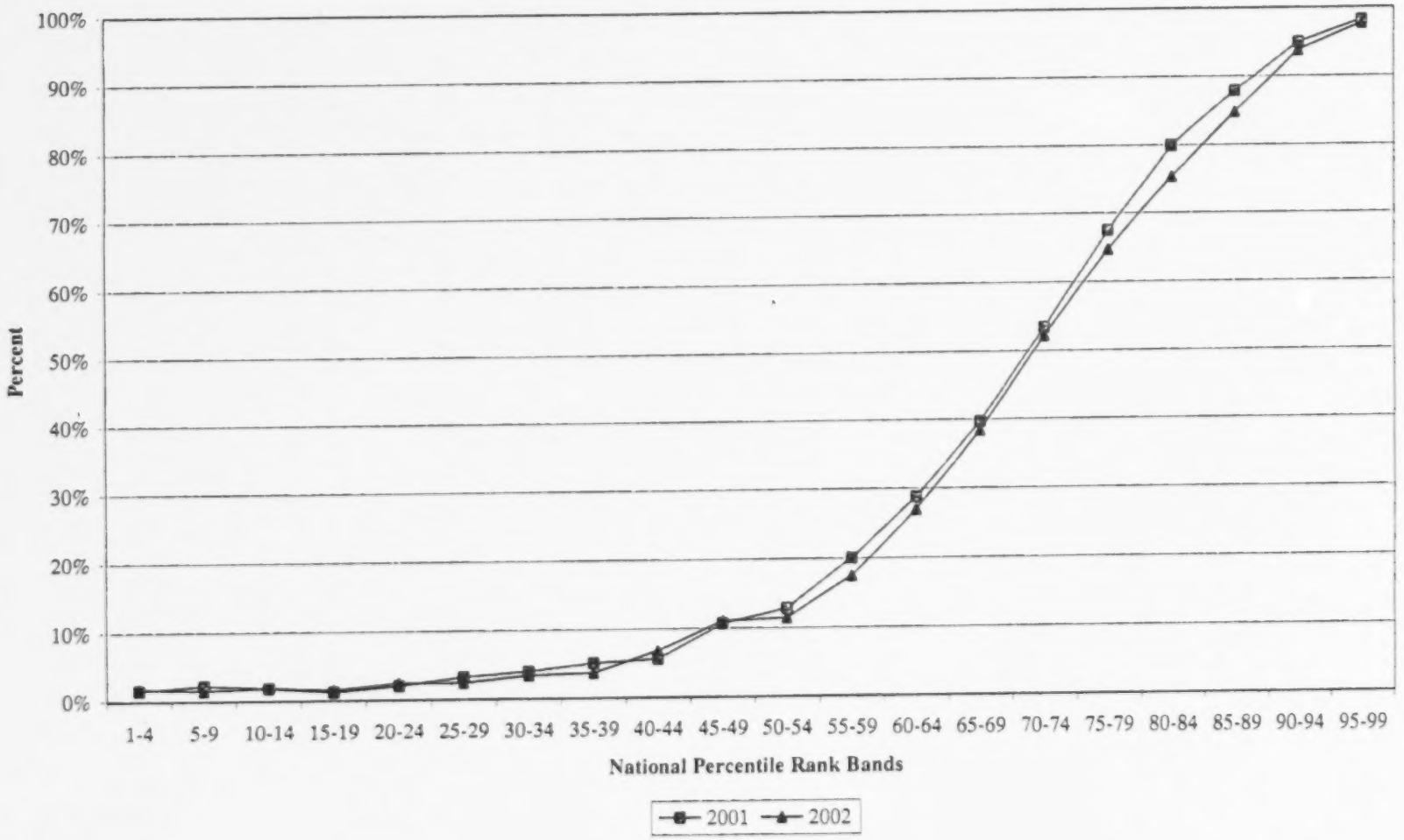


Figure 6. Percent of Students Scoring in 9th Grade ITED Math NPR Bands Subsequently Meeting 2001 and 2002 10th Grade WASL Math Standard



U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

ERIC Educational Resources Information Center

REPRODUCTION RELEASE

(Specific Document)

TM035031

I. DOCUMENT IDENTIFICATION:

Title: Validity Evidence for Washington Assessment of Student Learning (WASL) Performance Standard Cut-Scores for Reading and Mathematics	
Author(s): Duncan MacQuarrie	
Corporate Source:	Publication Date:

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

Level 1

X

Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) and paper copy

The sample sticker shown below will be affixed to all Level 2A documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2A

Level 2A

Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only

The sample sticker shown below will be affixed to all Level 2B documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2B

Level 2B

Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here, → please

Signature: Duncan MacQ		Printed Name/Position/Title: Duncan MacQuarrie, Manager of Student Assessment	
Organization/Address: Department of Research & Evaluation, Tacoma (WA) Public Schools, P.O. Box 1357, Tacoma, WA 98401-1357		Telephone: 253.571.1190	FAX: 253.571.1166
		E-Mail Address: dmacquar@tacoma.k12.wa.us	Date: 6/3/05

(Over)

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, or, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distnbutor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

ERIC CLEARINGHOUSE ON ASSESSMENT AND EVALUATION
UNIVERSITY OF MARYLAND
1129 SHRIVER LAB
COLLEGE PARK, MD 20742-5701
ATTN: ACQUISITIONS

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

ERIC Processing and Reference Facility
4483-A Forbes Boulevard
Lanham, Maryland 20706

Telephone: 301-552-4200
Toll Free: 800-799-3742
FAX: 301-552-4700
e-mail: ericfac@inet.ed.gov
WWW: http://ericfacility.org

EFF-088 (Rev. 2/2001)